

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 28, 2006, 12:23:55 ; Search time 779.148 Seconds
(without alignments)
3470.568 Million cell updates/sec

Title: US-10-511-270-2
Perfect score: 1702
Sequence: 1 MLGPQVWSSVRQGLSRSLR.....ELSPAEBAALMDFTSNGWL 327

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODBL=frame+ p2n.model -DEV=xlp
-Q=/abs/ABSSWES.spool/US10511270/runat_27022006_164254_20192/app_query.fasta_1
-DB=Published Applications NA Main -QFMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs05p
-USER=US10511270 -CGN 1 1 2064 @runat_27022006_164254_20192 -NCPU=6 -ICPU=3
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1696	99.6	2032	US-10-128-558-103	Sequence 103, App
2	1659	97.5	1061	US-10-511-270-1	Sequence 1, Appl1
3	1641.5	96.4	1521	US-10-120-988-156	Sequence 156, App
4	1497	88.0	1017	US-10-511-270-3	Sequence 3, Appl1
5	1430	84.0	1336	US-10-433-802-21	Sequence 21, Appl1
6	1402	82.4	1235	US-10-006-285-453	Sequence 453, App
7	1034	60.8	2581	US-10-108-260A-344	Sequence 344, App

8	727	42.7	664	US-10-424-599-136324	Sequence 136324,
9	720	42.3	456	US-10-128-558-317	Sequence 317, App
10	619	36.4	389	US-09-918-995-8256	Sequence 8256, Ap
11	605.5	35.6	54718	US-10-719-993-7038	Sequence 7038, Ap
12	570	33.5	867	US-10-369-493-33440	Sequence 33440, A
13	517.5	30.4	348	US-10-006-285-277	Sequence 277, App
14	382	22.4	942	US-10-369-493-42239	Sequence 42239, A
15	376.5	22.1	867	US-10-369-493-40233	Sequence 40233, A
16	375	22.0	13207	US-10-719-993-7016	Sequence 7016, Ap
17	375	22.0	23349	US-10-719-993-6841	Sequence 6841, Ap
18	336.5	19.8	245	US-10-006-285-111	Sequence 111, App
19	334.5	19.7	870	US-10-369-493-45137	Sequence 45137, A
20	331.5	19.5	933	US-10-369-493-39120	Sequence 39120, A
21	331	19.4	870	US-10-369-493-35026	Sequence 35026, A
22	329	19.3	897	US-10-369-493-31995	Sequence 31995, A
23	325.5	19.1	865	US-10-369-493-37079	Sequence 37079, A
24	325	19.1	945	US-10-156-761-6320	Sequence 6320, Ap
25	325	19.1	9025608	US-10-156-761-1	Sequence 1, Appl1
26	323.5	19.0	906	US-10-282-122A-39025	Sequence 39025, A
27	322.5	18.9	897	US-10-156-761-2503	Sequence 2503, Ap
28	322.5	18.9	925	US-10-369-493-36973	Sequence 36973, A
29	321	18.9	816	US-10-369-493-39488	Sequence 39488, A
30	320.5	18.8	852	US-10-369-493-24858	Sequence 24858, A
31	315.5	18.5	5069	US-09-070-927A-322	Sequence 322, App
32	313.5	18.4	873	US-10-214-556-1	Sequence 1, Appl1
33	313.5	18.4	873	US-10-073-741-1	Sequence 1, Appl1
34	312	18.3	870	US-10-369-493-24687	Sequence 24687, A
35	311	18.3	867	US-10-369-493-33837	Sequence 33837, A
36	310.5	18.2	879	US-10-369-493-41002	Sequence 41002, A
37	305	17.9	4737	US-10-238-075-1373	Sequence 1373, Ap
38	302	17.7	885	US-10-369-493-31553	Sequence 31553, A
39	301.5	17.7	750	US-09-974-300-509	Sequence 509, App
40	301.5	17.7	888	US-10-369-493-40937	Sequence 40937, A
41	300.5	17.7	873	US-10-369-493-37575	Sequence 37575, A
42	299	17.6	9170	US-10-650-274-67	Sequence 67, Appl1
43	298.5	17.5	894	US-10-282-122A-9404	Sequence 9404, Ap
44	297	17.5	882	US-10-238-075-1374	Sequence 1374, Ap
45	293.5	17.2	495269	US-10-398-221-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-10-128-558-103
Sequence 103, Application US/10128558
Publication No. US20040219521A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Wang, Zhiwei
APPLICANT: Weng, Gezh1
APPLICANT: Boyle, Bryan J
APPLICANT: Drmanac, Radoje T
TITLE OF INVENTION: Novel Nucleic Acids and
FILE REFERENCE: Polypeptides
CURRENT APPLICATION NUMBER: US/10/128,558
PRIOR FILING DATE: 2002-04-22
PRIOR APPLICATION NUMBER: US 60/339,453
PRIOR FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 09/488,725
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: US 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: PCT/US00/35017
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/491,404
PRIOR FILING DATE: 2000-01-25
PRIOR APPLICATION NUMBER: PCT/US01/02623
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: US 09/496,914
PRIOR FILING DATE: 2000-02-03
PRIOR APPLICATION NUMBER: US 09/560,875
PRIOR FILING DATE: 2000-04-27

102(e)
Seq.
103
60339453
2001 Prev.
12/11/200

```

; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05*
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 412
; SOFTWARE: pt_fl_genes Version 6.0
; SEQ ID NO 103
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (419)..(1402)
; US-10-128-558-103

```

Alignment Scores:	2.17e-193	length:	2032
Pred. No.:	1696.00	Matches:	326
Score:	99.7%	Conservative:	0
Percent Similarity:	99.7%	Mismatches:	1
Best Local Similarity:	99.6%	Indels:	0
Query Match:	8	Gaps:	0
DB:			

US-10-511-270-2 (1-327) x US-10-128-558-103 (1-2032)

Qy	1	MetLeuGlyProGlnValTrpSerSerValArgGlnGlyLeuSerArgSerLeuSerArg	20
Db	419	ATGCTGGGTCCCAAGTCTGTGTTCTGTGAGCAGGGGCTAAGCAGAGCTTGTCCAGG	478
Qy	21	AsnValGlyValTrpAlaSerGlyGlnGlyLysLysValAspIleAlaGlyIleTyrPro	40
Db	479	AATGTGGGGGTCTGGGCTCAGGGGAGGGGAAGAGTGACATTGCGGTATCTACCCC	538
Qy	41	ProValThrThrProPheThrAlaThrAlaGluValAspTyrGlyLysLeuGlnGluAsn	60
Db	539	CCTGTGACCACCCCTTCACTGCCACTGCAGAGGTGACTATGGAAACTGGAGAGAAT	598
Qy	61	LeuHisLysLeuGlyThrPheProPheArgGlyPheValValGlnGlySerAsnGlyGlu	80
Db	599	CTGCACAACTGGGCACCTTCCCTTCCGAGGCTTCGTGTCCAGGGCTCCATGGCGAG	658
Qy	81	PheProPheLeuThrSerSerGluArgLeuGluValValSerArgValArgGlnAlaMet	100
Db	659	TTTCCTTTCTGACCAGCAGTAGCGCTCGAGGTGTGAGCCGTGTGCCGACAGGCCATG	718
Qy	101	ProLysAsnArgLeuLeuLeuAlaGlySerGlyCysGlnSerThrGlnAlaThrValGlu	120
Db	719	CCCATGAACAGGCTCCTGCTAGCTGGCTCCGATGCGAGTCCACTCAAGCCACAGTGAAG	778
Qy	121	MetThrValSerMetAlaGlnValGlyAlaAspAlaAlaMetValValThrProCysTyr	140
Db	779	ATGACCGTCAGCATGGCCCAAGTTCGGGGCTGACGCGGCCATGTGTGTGACCCCTGTAC	838
Qy	141	TyrArgGlyArgMetSerSerAlaAlaLeuIleHisIleTyrThrLysValAlaAspLeu	160
Db	839	TATCGTGGCGCATGAGCAGTGGGCGCTCATTCACCACTACACCAAGTTGTGATCTC	898
Qy	161	SerProIleProValValLeuTyrSerValProAlaAsnThrGlyLeuAspLeuProVal	180
Db	899	TCTCCAATCCCTGTGTGTGTACATGTCCAGCCAAACAAGGGCTGACCTGCTGTG	958
Qy	181	AspAlaValValThrLeuSerGlnHisProAsnIleValGlyMetLysAspSerGlyGly	200
Db	959	GATGCAGTGGTCAACGCTTTCACAGCACCCGAATATTGTGGGCATGAAGACACAGCGGTGT	1018
Qy	201	AspValThrArgIleGlyLeuIleValHisLysThrArgLysGlnAspPheGlnValLeu	220
Db	1019	GATGTACCAAGATTGGGCTGATTGTTCACAAGACCAAGAGACGAGATTTTCAGGTGTG	1078
Qy	221	AlaGlySerAlaGlyPheLeuMetAlaSerTyrAlaLeuGlyValAlaValGlyValCys	240
Db	1079	GCTGATCGGCTGGCTTCTGATGCGCACGCTATGCCCTTGGAGCTGTGGGGGCGCTGTGC	1138

QY	241	AlAlLeuAlaasnValLeuGlYAlaGlnValCysGlnLeuGlubArgLeuCysCysThrGly	260
DB	1139	GGCCGTGGCCAATGTCCTGGGGGCTCAGGTGTGCCAGCTGGAGCCACTGTGTCACAGGGG	1198
QY	261	GlntRrPluaspAlaGlnLysLeuGlnHisArgLeuIleGluProasnAlaIaValThr	280
DB	1199	CAATGGGAAGATGCCCAGAAACTGCAGCACCGCCTCATTGAGCCAAACGCTGCGGTACC	1258
QY	281	ArgArgPheGlyIleProGlyLeuLysLysIleMetAspTrrPheGlyTyrTyrGlyGly	300
DB	1259	CGGCGCTTTGGGATCCACAGGGCTGAAGAAATCATGGACTGTTGGCTACTATGGAAGC	1318
QY	301	ProCysArgAlaProLeuGlnGluLeuSerProIaGluGluGluAlaLeuArgMetAsp	320
DB	1319	CCCTGCGCGCGCCCTTGACAGGAGCTGAGCCCGCTGAGGAGGAGGACACTGCCATGGAT	1378
QY	321	PheThrSerAenGlyTrrPleu	327
DB	1379	TTCAACGACAAACGGCTGGCTC	1399

QY 321 PhetHrSerAsnGlyTyrPleu 327
 |||||
 Db 1379 TTCACGACGACGCGCTC 1399

RESULT 2

US-10-511-270-1

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Alignment Scores:		
Pred. No.:	2.48e-189	1061
Score:	1659.00	326
Percent Similarity:	99.4%	
Best Local Similarity:	99.4%	1
Query Match:	97.5%	2
DB:	9	0

US-10-511-270-2 (1-327) x US-10-511-270-1 (1-1061)

QY	1 MetLeuGlyProGlnValTrpSerSerValArgGlnGlyLeuSerArgSerLeuSerArg	20
Db	8 ATGCTGGGTCCCAAGTCTGTCTTGTGAGGCAGGGGCTTAAGCAGAGCTGTCCAGG	67
QY	21 AsnValGlyValTrpAlaSerGlyGlnGlyLysLysValAspIleAlaGlyIleTyrPro	40
Db	68 AATGTGGGGCTCTGGGCTCAGGGGAGGGGAAGAAGGTGACATTGCCGGTATCTTA-CCC	126
QY	41 ProValThrThrProPheThrAlaThrAlaGluValAspTyr-GlyLysLeuGlnGluAs	60
Db	127 CCTGTGACCAACCCCTTCACTGCCACTGCAGAGGTGACTATGGGGAAC TGAGAGAGAA	186
QY	60 nLeuHisLysLeuGlyThrPheProPheArgGlyPheValValGlnGlySerAsnGlyGly	80
Db	187 TCTGCACAAACTGGGCACTTCCCTTCCGAGGCTTCTGTGTCAGGGCTCCCAATGGCGA	246

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - nucleic search,

Run on: February 28, 2006, 11:51:51 ; Search time 6074.23 Seconds

(without alignments)
3060.111 Million cell updates/sec

Title: US-10-511-270-2
Perfect score: 1702
Sequence: 1 MLGPQWSSVRQGLSRLSR.....ELSPAREEARLMDPTSGWL 327

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/abs/ABSSWEB.spool/US10511270/runat_27022006_164246_19980/app_query.fasta_1
-DB=GenEmbl -QFWT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCH=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs03h
-USER=US10511270 @CGN 1 1 5142 @runat_27022006_164246_19980 -NCPU=6 -ICPU=3
-NO MMAP -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Match	Length	DB	ID	Description
1	1702	100.0	2480	8	BC045550	BC045550 Homo sapi	BC045550 Homo sapi
2	1641.5	96.4	1521	6	AR541908	AR541908 Sequence	AR541908 Sequence
3	1516	89.1	1618	9	BC016430	BC016430 Mus muscu	BC016430 Mus muscu

4	1441.5	84.7	1442	8	HSA312051	AJ312051 Homo sapi
5	1430	84.0	1336	6	AX458350	AX458350 Sequence
6	1418.5	83.3	884	6	CQ722087	CQ722087 Sequence
7	1182.5	69.5	1148	5	BC087798	BC087798 Xenopus t
8	1172	68.9	1160	5	BC084421	BC084421 Xenopus 1
9	1155	67.9	1370	5	BC066708	BC066708 Danio rer
10	1034	60.8	2581	6	AX833220	AX833220 Sequence
11	1034	60.8	2581	8	AK094791	AK094791 Homo sapi
12	791.5	46.5	1978	8	BC057821	BC057821 Homo sapi
13	791.5	46.5	2012	8	BC011916	BC011916 Homo sapi
14	755	44.4	933	5	CR407189	CR407189 Gallus ga
15	738	43.4	505	6	BD111539	BD111539 EST and e
16	738	43.4	505	6	AR415986	AR415986 Sequence
17	738	43.4	505	6	AX972820	AX972820 Sequence
18	605.5	35.6	157193	8	AL355315	AL355315 Human DNA
19	568	33.4	265464	14	AC157384	AC157384 Bos tauru
20	564.5	33.2	149854	9	AC117225	AC117225 Mus muscu
21	564.5	33.2	161150	9	AC133503	AC133503 Mus muscu
22	564.5	33.2	223987	14	AC147800	AC147800 Mus muscu
23	536.5	31.5	133470	9	AC131867	AC131867 Rattus no
24	536.5	31.5	222812	14	AC158177	AC158177 Rattus no
25	394.5	23.2	110000	1	AY596297_01	Continuation (2 of
26	382	22.4	16155	1	AB004999	AB004999 Halobacte
27	375.5	22.1	110000	1	AB017225	Continuation (37 o
28	375.5	22.1	110000	1	AB017334	Continuation (37 o
29	375.5	22.1	110000	1	AB017355	Continuation (37 o
30	375.5	22.1	110000	1	CP000001	Continuation (37 o
31	375.5	22.1	291030	1	AB017036	Continuation (37 o
32	375.5	22.1	302481	1	AB017010	AB017010 Bacillus
33	375	22.0	14300	8	HSA304804	AB017010 Homo sapi
34	370	21.7	909	6	AR347015	AR347015 Sequence
35	365	21.4	110000	1	AB017333	Continuation (34 o
36	365	21.4	110000	1	CP000002	Continuation (34 o
37	361.5	21.2	110000	1	AB017333	Continuation (36 o
38	361.5	21.2	110000	1	CP000002	Continuation (36 o
39	356	20.9	5240	6	AR310393	AR310393 Sequence
40	356	20.9	5240	15	AF056977	AF056977 Penicilli
41	355.5	20.9	287765	1	AB017276	AB017276 Bacillus
42	355	20.9	45055	14	AY714867	AY714867 Unculture
43	353	20.7	11714	1	AB013350	AB013350 Methanosa
44	352	20.7	4183	1	AF319635	AF319635 Thermococ
45	347.5	20.4	110000	1	AP008226_06	Continuation (7 of

ALIGNMENTS

RESULT 1	BC045550	2480 bp	mRNA	linear	PRI 28-JUL-2005
LOCUS	BC045550				
DEFINITION	Homo sapiens chromosome 10 open reading frame 65, mRNA (cDNA clone				
ACCESSION	BC045550				
VERSION	BC045550.1	GI:28278549			
KEYWORDS	MGC.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 2480)				
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ugin, T.B., Toshiki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hultky, S.W., Villalón, D.R., Muzny, D.M., Sodergren, B.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,				

102(a)
JOURNAL PUBMED REFERENCE AUTHORS TITLE JOURNAL
JOURNAL TITLE JOURNAL
REMARK COMMENT
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.B., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A. Mammalian Gene Collection Program Team Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002) • 12477932
2 (bases 1 to 2480)
Dec. 2002

NIH MGC Project
Direct Submission
Submitted (31-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-rc@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAX Plate: 106 Row: 1 Column: 3
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 31543059.
Location/Qualifiers
1. 2480

FEATURES
source
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:57219 IMAGE:5274464"
/issue_type="Brain, hippocampus"
/clone_lib="NIH MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescriptR"
1. 2480
/gene="C10orf65"
/db_xref="GeneID:112817"
344. 1327
/gene="C10orf65"
/codon_start=1
/product="hypothetical protein LOC112817"
/protein_id="AAH45550.1"
/db_xref="GI:28278550"
/translation="MGPOVWSSVRQGLSRNNGVWASGEKKVDIAGIYPPVTT
PFTATAEVDYKLEENLHKLGPFRGFVVOGNGEPPLTSSERLEVSRRVQAMPK
NRLLAGSGCESTQATVEMTVMAQVADAAVVTPTCYRGRMSAALIHHTYKADL
SPIPVLYSVANTGLDLPDAVVTLSHPNIVGMKDSGDVTRIGLIHKTQKDFQ
VLASAGFIASVYALGAVGVCALANVLGAQVCQLERLCTGOWEDAOKLOHRLLEPN
AAVTRRFGIPGLKIMDFWFGYGGPCRAPLQELSPAEERLRMDFTSNGWL"

ORIGIN
Alignment Scores:
Pred. No.: 4.34e-124 Length: 2480
Score: 1702.00 Matches: 327
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 100.0% Indels: 0
DB: 8 Gaps: 0
us-10-511-270-2 (1-327) x BC045550 (1-2480)
1 MetLeuGlyProGlnValTrpSerSerValArgGlnGlyLeuSerArgSerLeuSerArg 20

|||||
Db 344 ATCTGGTCCCAAGTCTGTCTTCTGTAGGCGAGGGCTAAGCAGAGCTTTCAGG 403
QY 21 AsnValGlyValTrpAlaSerGlyGlyValysValAspIleAlaGlyIleTyrPro 40
Db 404 AATGTGGGGTCTGGGCTCAGGGAGGAGAGAGAGTGCATTGCGGTATTCACCC 463
QY 41 ProValThrThrProPheThrAlaThrAlaGluValAspTyrGlyValysLeuGluAsn 60
Db 464 CCTGTGACCAACCCCTTCACTGACCACTGCAAGGTGACTATGGAACTGAGAGAAAT 523
QY 61 LeuHisIleLeuGlyThrPheProPheArgGlyPheValValGlnGlySerAsnGlyGlu 80
Db 524 CTGCACAACTGGGCACTTCCCTTCGAGGCTTCGTGTCCAGGGCTCCAATGGCGAG 583
QY 81 PheProPheLeuThrSerSerGluArgLeuGluValValSerArgValArgGlnAlaMet 100
Db 584 TTTCCTTCTGACCAAGCAGTGAAGCGCTCGAGGTGTGAGCGCGTGGCCAGGCATG 643
QY 101 ProLysAsnArgLeuLeuLeuAlaGlySerGlyCysGluSerThrGlnAlaThrValGlu 120
Db 644 CCCAAGAACAGGCTCTGCTAGCTGGGCTCCGGATGCGAGTCCACTCAAGCCACAGTGA 703
QY 121 MetThrValSerMetAlaGlnValGlyAlaAspAlaAlaMetValThrProCysTyr 140
Db 704 ATGACCGTCAGCATGGCCAGGTCGGGCTGACGCGGCCATGCTGTGACCCCTGTCTAC 763
QY 141 TyrArgGlyArgMetSerSerAlaAlaLeuIleHisIleTyrThrIysValAlaAspLeu 160
Db 764 TATGTGGCCGATGAGCAGTGGCGCCCTCATTCACCACTACCAAGGTGTGATCTC 823
QY 161 SerProIleProValValLeuTyrSerValProAlaAsnThrGlyLeuAspLeuProVal 180
Db 824 TCTCAATCCCTGTGTGTGTGTACAGTGTCCAGCAACAGGCGTGAACCTGCTGTG 883
QY 181 AspAlaValValThrLeuSerGlnHisProAsnIleValGlyMetLysAspSerGlyGly 200
Db 884 GATGACGTGTACGCTTTCACAGCACCAGATATGTGTGGCATGAGGACGCGGTGT 943
QY 201 AspValThrArgIleGlyLeuIleValHisIleThrArgLysGlnAspPheGlnValLeu 220
Db 944 GATGTACCAAGATTGGGCTGATTGTTCACAGACCAAGAGCAGGATTTCAGGTGTG 1003
QY 221 AlaGlySerAlaGlyPheLeuMetAlaSerTyrAlaLeuGlyAlaValGlyValCys 240
Db 1004 GCTGATCGGCTGCTTCTGTATGCGCACTATGCTTGGAGCTGTGGGGGGCTGTGC 1063
QY 241 AlaLeuAlaAsnValLeuGlyAlaGlnValCysGlnLeuGluArgLeuCysCysThrGly 260
Db 1064 GCCCTGGCCATGTCTTGGGGCTCAGGTGTGCCAGCTGAGCGACTGTGCACGGGG 1123
QY 261 GlnTrpGluAspAlaGlnIleLeuHisIleArgLeuIleGluProAsnAlaAlaValThr 280
Db 1124 CAATGGAGAGATGCCAGAACTGACAGCAGCGCTCATGACCAACGCTGCGGTGACC 1183
QY 281 ArgArgPheGlyIleProGlyLeuIleValysIleMetAspTrpPheGlyTyrTyrGlyGly 300
Db 1184 CGGCGTTTGGGATCCCAAGGCTGAAGAAATCATGACTGTTTGGCTACTATGAGGC 1243
QY 301 ProCysArgAlaProLeuGlnIleLeuSerProAlaGlnGluAlaLeuArgMetAsp 320
Db 1244 CCTGCGCCGACCCCTTGCAGAGCTGAGCCCGCTGAGAGAGAGGCACTGCGCATGAT 1303
QY 321 PheThrSerAsnGlyTyrLeu 327
Db 1304 TTCACCAAGCAAGGCTGCTC 1324

102(e)

RESULT 2
AR541908 1521 bp DNA linear PAT 08-OCT-2004
LOCUS AR541908
DEFINITION Sequence 156 from patent US 6743619.
ACCESSION AR541908
VERSION AR541908.1 GI:539333988

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 09:41:19 ; Search time 4837.81 Seconds
(without alignments)
10626.906 Million cell updates/sec

Title: US-10-511-270-1
Perfect score: 1061
Sequence: 1 gaagtcctatgctggtccccc.....gcacttcgagcctgaattcc 1061

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank:
1: gb_ha:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1025.4	96.6	2480	9 BC045550	BC045550 Homo sapi
2	818.6	77.2	1442	9 HSA312051	AJ312051 Homo sapi
3	814.2	76.7	1336	6 AX458350	AX458350 Sequence
4	782.4	73.7	1521	6 AR541908	AR541908 Sequence
5	761	71.7	884	6 CQ722087	CQ722087 Sequence
6	722.8	68.1	1618	10 BC016430	BC016430 Mus muscu
7	445.6	42.0	2581	6 AX833320	AX833320 Sequence
8	445.6	42.0	2581	9 AK094791	AK094791 Homo sapi
9	422.6	39.8	1160	5 BC084421	BC084421 Xenopus l
10	405	38.2	1370	5 BC066708	BC066708 Danio rer
11	351	33.1	2012	9 BC011916	BC011916 Homo sapi
12	349.4	32.9	1978	5 BC057821	BC057821 Homo sapi
13	329.4	31.0	933	5 CR407189	CR407189 Gallus ga
14	317.2	29.9	505	6 AR415986	AR415986 Sequence
15	317.2	29.9	505	6 AX972820	AX972820 Sequence
16	317.2	29.9	505	6 BD111539	BD111539 EST and e
17	217	20.5	157193	9 AL355315	AL355315 Human DNA
18	189	17.8	14300	9 HSA304804	AJ304804 Homo sapi
19	137.6	13.0	149854	10 AC117225	AC117225 Mus muscu

C 20	137.6	13.0	161150	10 AC133503	AC133503 Mus muscu
C 21	137.6	13.0	223987	2 AC147800	AC147800 Mus muscu
C 22	129.4	12.2	133470	10 AC131867	AC131867 Rattus no
C 23	123.8	11.7	16155	1 AB004999	AB004999 Halobacte
C 24	112.6	10.6	458	11 G80217	G80217 S208P6433FC
C 25	95.4	9.0	110000	1 AB017180_01	Continuation (2 of
C 26	89.2	8.4	10722	1 AB011168	AB011168 Methanosa
C 27	85.6	8.1	346362	1 BX640439	BX640439 Bordetell
C 28	85.6	8.1	348257	1 BX640425	BX640425 Bordetell
C 29	81.2	7.7	11714	1 AB013350	AB013350 Methanosa
C 30	79.6	7.5	879	6 CQ878055	CQ878055 Sequence
C 31	79.6	7.5	12967	1 AB004533	AB004533 Pseudomon
C 32	78	7.4	876	6 CQ878057	CQ878057 Sequence
C 33	75	7.1	11371	1 AE012275	AE012275 Xanthomon
C 34	74.4	7.0	5240	6 AR310393	AR310393 Sequence
C 35	74.4	7.0	5240	8 AF056977	AF056977 Penicillii
C 36	73.6	6.9	348580	1 BX572601	BX572601 Rhodopseu
C 37	71.2	6.7	10293	1 AE011807	AE011807 Xanthomon
C 38	70.6	6.7	301925	1 AP005046	AP005046 Streptomy
C 39	70.4	6.6	272101	1 AE017302	AE017302 Thermus t
C 40	69	6.5	346879	1 BX572598	BX572598 Rhodopseu
C 41	68.6	6.5	283100	1 SC0939110	SC0939110 Streptomy
C 42	68.6	6.5	283100	1 AP006618_40	Continuation (41 o
C 43	68.4	6.4	110000	1 AB017333_35	Continuation (36 o
C 44	66.4	6.3	110000	1 CP000002_35	Continuation (36 o
C 45	66.4	6.3	110000	1	

ALIGNMENTS

RESULT 1
LOCUS BC045550
DEFINITION Homo sapiens chromosome 10 open reading frame 65, mRNA (CDNA clone
MGC:57219 IMAGE:5274464), complete cds.
VERSION BC045550
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.P., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Ustin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.B.,
Scherer, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE
JOURNAL PUBMED
AUTHORS
REFERENCE
JOURNAL
Submitted (31-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

Dec. 11,